

## HELP SHEET 1

You decide to ask a graduate student from a lab that one of your group members is working in, “How do we apply Hardy-Weinberg to two populations?” The graduate student points out some additional ramifications of the formula you have correctly used to interpret the data from Site 1. That is, the expected frequency of heterozygotes, assuming samples are drawn from a single population, i.e., without population structure, is:

$$2pq$$

Although selection can play an important role, population subdivision is also driven by genetic drift. As two subpopulations stop interacting, random changes in allele frequency due to genetic drift cause the two subpopulations to diverge genetically.

As allele frequencies of populations drift away from each other, the number of expected heterozygotes should change. However, if population structuring is not known and subpopulations are treated as a single population, the observed number of heterozygotes decreases compared to expectations under the Hardy-Weinberg Equilibrium (HWE).

$$\frac{2p_1q_1 + 2p_2q_2}{2}$$

The graduate student reminds you that  $1 - p = q$ . However, if the population is evenly subdivided, as evident from different allele frequencies, then the expected frequency of heterozygotes is found by averaging the two (sub)populations: The graduate student explains that population subdivision can also be thought of as inbreeding (an excess of homozygotes because of mating-related individuals): population subdivision decreases the sizes of subpopulations relative to the whole population. Smaller population size increases the probability of mating between relatives, even when mating is random within subpopulations. When mating occurs between even distantly related individuals that share the same alleles, their offspring are more likely to be homozygous than when mating occurs randomly in larger populations. The net result is a decrease in the frequency of heterozygotes. The expected frequency of heterozygotes found by averaging the two subpopulations, which is intended to account for population structure, will be smaller than the expected frequency of heterozygotes determined without assuming population structure ( $2pq$  for the entire population) if allele frequencies differ between subpopulations, ( $p_1 \neq p_2$ ) i.e., populations are subdivided. Thus a decrease in the heterozygosity may be indicative of population subdivision.



## HELP SHEET 2

She explains that calculating heterozygote frequencies in subdivided populations may not be quite as simple as she had previously indicated. She explains that the term the *Wahlund effect* is used for the decrease in expected heterozygote frequency when sampling subdivided populations relative to treating them as a whole population. This Wahlund effect is the name for what she was explaining earlier and the distinction it represents should be evident in your calculations in 4.1. This phenomenon is actually used by population geneticists to determine if population subdivision is occurring. The significance of this difference can be validated statistically using  $F$  statistics that are often used to measure subdivision and describes the expected level of heterozygosity.

They were developed by Sewall Wright and are often called “inbreeding coefficients” because they are based on deficiencies of heterozygotes. Population subdivision decreases the sizes of subpopulations relative to the whole population. Subdivided populations thus not only undergo genetic drift independently of each other (without migration), but also at a faster rate because of decreased population sizes. Additionally, this subdivision increases the probability of mating between relatives, even when mating is random within subpopulations. When mating occurs between individuals that share the same alleles because they are related, their offspring are more likely to be homozygous than if mating occurs randomly in larger populations.  $F_{ST}$  is probably the most commonly used  $F$  statistic given its value in looking explicitly at population subdivision.

$$F_{ST} = 1 - \frac{H_S}{H_T} = 1 - \frac{\text{(Heterozygosity expected within subpopulations)}}{\text{(Expected heterozygosity across total population)}}$$

The heterozygosity expected within subpopulations is based on allele frequencies in subpopulations. That is, the expected heterozygosity within subpopulations ( $H_S$ ) can also be expressed as the sum of the number of individuals in a subpopulation divided by the number in the total population, multiplied by the expected frequency of heterozygotes under HWE based on observed allele frequencies in each subpopulation. This is a weighted average for the two subpopulations; the weighting is by sample size. For an area that is subdivided into two subpopulations:

$$H_S = \frac{n_1}{N} * 2p_1q_1 + \frac{n_2}{N} * 2p_2q_2$$

where  $n_1$  and  $n_2$  are the sample sizes for subpopulations 1 and 2, respectively, and  $N$  is the total sample size. Thus,  $F_{ST}$  weights differences in allele frequencies by the size of the subpopulation. The more allele frequencies diverge from the original population due to subdivision, the larger the  $F_{ST}$  value will be. Furthermore, larger subpopulations that diverge from the total population will have a larger impact on  $F_{ST}$  than smaller, equally divergent, subpopulations.

There are many examples in the literature of researchers using a variety of molecular markers to calculate  $F_{ST}$ . This statistic can range from 0–1, with 0 indicating complete random mating (panmixis) and 1 indicating completely structured populations. Wright indicated that interpretation of  $F_{ST}$  can be difficult, as even small  $F_{ST}$  values can be indicative of differentiation. However, Wright suggested guidelines for interpreting  $F_{ST}$  values. We will use these guidelines for interpreting  $F_{ST}$  values in this case study. Wright’s guidelines for interpreting  $F_{ST}$  values are:

- 0–0.05 may indicate little genetic differentiation.
- 0.05–0.15 indicates moderate genetic differentiation.
- 0.15–0.25 indicates great genetic differentiation.
- >0.25 indicates very great genetic differentiation.

